Data & Policy (2025), 7: e25
doi:10.1017/dap.2024.49

DATA FOR POLICY REPORT

Data technologies and analytics for policy and governance: a landscape review

Omar Isaac Asensio[1,2], Catherine E. Moore[1], Nicola Ulibarri[3], Mecit Can Emre Simsekler[4,5], Tian Lan[6,7] and Gonzalo Rivero[8]

[1]School of Public Policy, Georgia Institute of Technology, Atlanta, USA
[2]Institute for Data Engineering & Science (IDEaS), Georgia Institute of Technology, Atlanta, USA
[3]Department of Urban Planning and Public Policy, University of California Irvine, Irvine, USA
[4]Department of Management Science and Engineering, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates
[5]School of Management, University College London, London, UK
[6]Department of Geography, University College London, London, UK
[7]School of Resource and Environmental Sciences, Wuhan University, Wuhan, China
[8]Independent Researcher, Washington, DC, USA
Corresponding author: Omar Isaac Asensio; Email: asensio@gatech.edu

Received: 24 February 2022; **Revised:** 04 April 2024; **Accepted:** 24 July 2024

Keywords: digital data; digital twins; distributed ledger systems; internet of things; machine learning and artificial intelligence

Abstract

Data for Policy (dataforpolicy.org), a trans-disciplinary community of research and practice, has emerged around the application and evaluation of data technologies and analytics for policy and governance. Research in this area has involved cross-sector collaborations, but the areas of emphasis have previously been unclear. Within the Data for Policy framework of six focus areas, this report offers a landscape review of Focus Area 2: Technologies and Analytics. Taking stock of recent advancements and challenges can help shape research priorities for this community. We highlight four commonly used technologies for prediction and inference that leverage datasets from the digital environment: machine learning (ML) and artificial intelligence systems, the internet-of-things, digital twins, and distributed ledger systems. We review innovations in research evaluation and discuss future directions for policy decision-making.

Policy Significance Statement

A growing and robust community is deploying technologies and analytics to address public policy challenges. This landscape review highlights historical trends and priority areas for Data for Policy Area 2: Technologies and Analytics. We review characteristics of submissions from academic and nonacademic authors, comment on relationships from data collection to decision-making, and document advances in policy analytics related to ML, the internet-of-things, digital twins, and distributed ledger system technologies.

1. Introduction

Since 2015, *Data for Policy* has established itself as a leading global forum for cross-sectoral and interdisciplinary exchange on digital revolutions in policy-making (Verhulst et al., 2019). The conference

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

and its partner journal, Data and Policy, serve a diverse community network that spans a range of disciplines and sectors. The journal has organized this field into six focus areas to capture the emerging trends shaping global discussion (Engin et al., 2024). Here, we present a report for Focus Area 2: Technologies and Analytics, which expands on both established and new data streams from personal, proprietary, administrative, and public sources and surveys the current landscape of analytical technologies and challenges for both practitioners and researchers.

A community of research and practice has emerged around the deployment of data processing technologies and analytical tools for evidence-based policy-making (Kim et al., 2014; Suominen and Hajikhani, 2021). This community is increasingly using data analytics to generate field evidence and large-scale case studies to inform policymaking on various societal challenges (Verhulst et al., 2019; Mergel et al., 2016; Anshari et al., 2018). Drawing on foundational knowledge from computational social science (CSS) (Lazer et al., 2020) and related fields at the data science-policy interface, this community includes collaborators from academia, industry, and government. As a rapidly evolving field, the collection of technologies and platforms has a breadth of social implications given that technology developments occur more rapidly than use regulations can be established. In this landscape review, we discuss policy data interactions and topics established within *Data for Policy* over the last few years.

We highlight three fundamental challenges for the implementation of data tools within policy analysis. First, a greater focus on causality is needed where the objective is to uncover the effects of policy changes on a given population. We have seen increased adoption of experimental methods (such as randomized controlled trials) in policy studies, but their use still lags behind traditional approaches (Angrist & Pischke, 2010). With access to larger datasets, especially in the context of digital trace data, we expect a greater emphasis on new methods that can be combined with machine learning (ML) to disentangle causality in observational datasets with potentially many more variables than observations (Athey & Imbens, 2017). These integrative approaches can also offer valuable insights into understanding heterogeneity, which can bring us closer to estimating individual causal effects—and meaningfully distinguish between personalized and population-based decision-making (Athey & Imbens, 2016; Mueller and Pearl, 2023; Wager & Athey, 2018). To overcome challenges in applying causal theories or targeting policy interventions, the *Data for Policy* community is increasingly encouraging counterfactual thinking, especially by leveraging a combination of both experimental and observational data.

Second, to translate insights from massive quantities of data, the community is increasingly engaging specialists in cross-sector collaborations. This includes data scientists and managers in public or private organizations who have decision authority in data collection platforms and governance, as well as participatory research activities engaging communities affected by policy or technology decisions. For example, of Data for Policy Conference Focus Area 2 submissions for 2021 and 2022, one-fourth of the 129 authors represented cross-sector authorship (e.g. academic-industry, academic-government, academic-non-governmental organizations (NGOs)) as is depicted in Figure 1. We argue that this scientific model of collaborative research is important to accelerate translational research, provide testbeds for learning, develop use cases for data innovations, and meaningfully bridge competing cultures between research and practice.

Third, considering the rate of data innovations in the private sector by platform owners, aggregators, and intermediaries, there is a constant interplay between increasing data access (and therefore greater capabilities to analyze human behavior) and increasing data protections. A classification of submitted Area 2 abstracts to *Data for Policy 2021* revealed the most relevant concepts to be data mining, big data analytics, and data protection challenges, which is consistent with this relationship. These keywords were derived directly from the submitted user abstracts and were not predetermined by the journal.

Figure 2 conceptualizes this typical scaffolding between the data, the analytical tools and technologies used, and the applications that support evidence-based decision-making. The middle layer, which represents Technologies and Analytics, is built upon increasingly complex data infrastructure and is constantly evolving. Specifically, this paper expands upon ML, the Internet of Things (IoT), digital twins, and blockchain and distributed ledger systems (DLSs), but the layer encompasses additional tools that are characteristic of Data for Policy's Focus Area 2 (Engin et al., 2024) and leaves room for the continual evolution of eligible technologies.

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

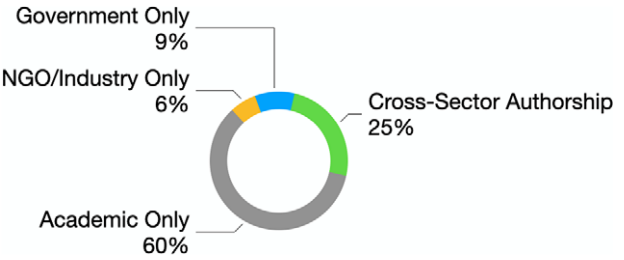


Figure 1. *Data for Policy Contributing Authors for Focus Area 2: Technologies and Analytics (2021–2022) The majority (60 percent) of the 129 submitting authors were solely from academic institutions, while one-fourth represented cross-sector authorship (academic-government, academic-industry, and/or academic-NGO scientific collaborations), 9 percent were government authors, and 6 percent were NGO/Industry authors only.*

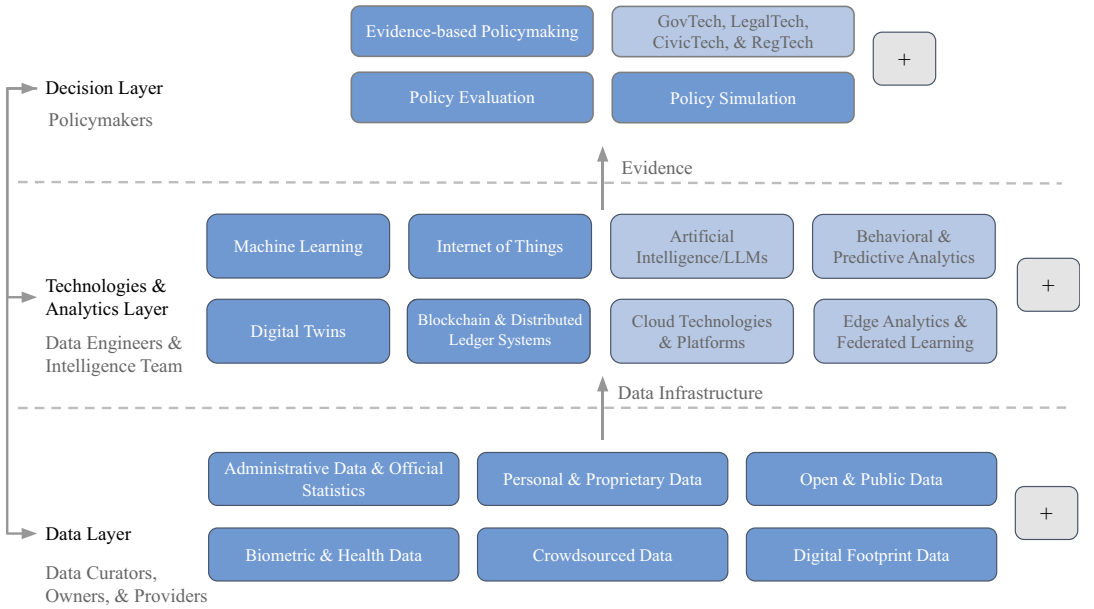


Figure 2. *From Data to Decision-Making: A Conceptual Framework of Data-Policy Interactions within Focus Area 2: Technologies and Analytics. The darker shaded boxes indicate topics covered in this Landscape Review. The lighter-shaded boxes indicate topics that exist under the Focus Area 2 umbrella but are not included in this paper.*

The remainder of this report is organized as follows: In Section 2, we review aspects of modern research data collection, associated information and real-time communication technologies, and government and administrative records. Section 3 reflects on four data tools that uniquely leverage data from the digital environment: ML applied to policy decision-making (Section 3.1); the internet-of-things in smart, connected infrastructure (Section 3.2); digital twins technologies for planning and design in the built environment (Section 3.3); and distributed ledger technologies that capture distributed trust (Section 3.4). In Section 4, we comment on the application of these analytical tools for policy evaluation.

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

2. Data sources

Policymakers have always had access to a variety of conventional data sources for measurement and evaluation in government statistics, often in the form of surveys that attempt to measure population trends, health, education, crime, and other aspects of social life (Groves, 2011). However, high-quality population surveys are complicated logistical operations that can be expensive or infrequently conducted. The gradual decline in response rates over recent decades (Brick and Williams, 2013; Singer, 2006) has motivated the search for alternative sources of data. This wealth of new digital data sources has led to the development of technical solutions for the storage, manipulation, and analysis of data (Lazer and Radford, 2017).

2.1. Conventional data sources

Government and administrative records have been valuable as a form of big data in social science research (Connelly et al., 2016). Administrative data, collected primarily for nonstatistical purposes, is readily available to governments and can be used to produce estimates of attributes that are not easily captured in surveys (Nordbotten, 2010). An important element in this discussion is the potential value of record linkage which allows for the integration of different data streams—for instance, combining census responses with tax and property data. Despite these advantages to overcome data silos, there is growing concern over incidental disclosure and reidentification. Electronic health records and other personally identifiable data have captured attention in this context, especially regarding the improvement and personalization of care (Cebul et al., 2011; Cowie et al., 2017; Abul-Husn and Kenny, 2019). In addition to privacy concerns, survey data is infrequently updated, which limits our ability to measure social phenomena and evaluate the impacts of policy changes. For example, in the United States, measures of household energy consumption are updated within the Residential Energy Consumption Survey 3 (RECS) and Commercial Buildings Energy Consumption Survey (CBECS), which are conducted every few years. As a result, monitoring the effects of energy-efficient and emissions-reducing policies becomes impractical for impact evaluation (EIA, 2022).

2.2. Digital data sources

Digital innovations in nearly every aspect of modern life have had a transformative impact on the availability of data that can be used to study and inform public policy (Salganik, 2019; Jungherr et al., 2020). Every day, a large portion of our behaviors are captured by digital systems (Golder and Macy, 2014). This is true for the increasing number of activities that are mediated by a computer or a cellular device (ranging from web browsing, connecting on social media, and utilizing mobile apps). Similar arguments can be made for the myriad of sensors and electronic systems with which we interact daily. Other examples include public or private CCTV systems (Taylor and Gill, 2014) and a wide array of sensors in public and private spaces (Ratti and Claudel, 2016; van De Sanden et al., 2019).

The decentralized nature of digital data collection and non-representative datasets allows for the repurposing of information for secondary uses (Salganik, 2019). For example, in urban analytics, smart cities projects (Batty et al., 2012; Albino et al., 2015) have taken advantage of active and passive monitoring devices (Singleton et al., 2017; Asensio et al., 2021) to study mobility (Aguilera and Boutueil, 2018), criminality (Ferguson, 2017; Meijer and Wessels, 2019), and the resilience of urban infrastructure in natural disasters (Khalaf et al., 2015; Dong and Shan, 2013). Further, mobile devices such as smartphones and wearable technology (Gandy et al., 2017) have become valuable sources of data for automated contract tracing and data collection. These devices consistently capture metrics related to browsing behavior, geolocation (Nikolic and Bierlaire, 2017; Wu et al., 2013), patterns of communication (Green et al., 2021; Blumenstock et al., 2015), and other features that can be used to study real-time aspects of human behavior.

Websites and social media also serve as valuable sources of digital data. Individuals and organizations maintain an online presence for a large variety of activities—learning, shopping, dating, and so forth

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

Some of these activities occur in public spaces into which researchers can gain programmatic access via structured (for instance, through a REST API) or unstructured methods (web scraping or web harvesting); or they can be investigated by the providers of the service through internal data (Kramer et al., 2014; Yang et al., 2021). This has allowed for aspects of public and private life to be studied at scale. For example, internet and social media data have played a role in the research on public attention, communication, and public health (Klašnja et al., 2015; Dugas et al., 2013; Arora et al., 2019). Additionally, satellite and aerial photography have been widely used for the analysis of urban and rural environments, including studying patterns of growth, mobility, and poverty (Wania et al., 2014; Zhang et al., 2019; Jean et al., 2016). Images from interactive panoramas like Google Street View have been used to gain insights into subjective perceptions of streetscapes (Ye et al., 2019; Liu et al., 2019; Rundle et al., 2011; Liu et al., 2017). For further discussion on the alternative applications of data from the digital environment to address public issues, see Verhulst (2021).

2.3. Challenges and opportunities underlying data sources

The existing literature characterizes digital data as advantageous compared to surveys, which have limitations that threaten the accuracy of data collection (Salganik, 2019). For example, due to cognitive biases, recalling past behavior is burdensome for many respondents. This limits the accuracy of survey responses especially as respondents are asked to recall information over longer periods of time (Grotpeter, 2008). Additionally, respondents are less likely to truthfully report answers to sensitive questions when there is an interviewer present (Tourangeau and Yan, 2007; Krumpal, 2013). Although methods exist to address these shortcomings (Lensvelt-Mulders et al., 2005; Blair and Imai, 2012), the ability to directly measure behaviors offers advantages in the availability, accuracy, and depth of the data. This is clearly illustrated in the use of GPS-enabled devices to complement mobility data from large transportation surveys (Stopher et al., 2007). Rather than asking a small sample of respondents to recall detailed information, digital data offers the opportunity to study a larger population, with greater accuracy, precision, and detail (see, for instance, (Merry and Bettinger, 2019; Wolf et al., 2003)).

Digital data also poses several relevant challenges. For example, in the study of public opinion, social media can be considered an attractive data source for its potential to replace traditional polls (Tumasjan et al., 2011; McKelvey et al., 2014; DiGrazia et al., 2013). However, not everyone owns a digital device, has access to the internet, or maintains a social media presence. This means that the pool of individuals who can be studied is not always representative of the general public. This digital divide has been documented among adults and the right to vote (Gayo-Avello et al., 2011; Murphy et al., 2014; Barberá and Rivero, 2014). In an effort to mitigate biases in population sampling, suitable weighting methods have emerged (Sen et al., 2019; Elliott and Valliant, 2017; Kennedy et al., 2016). Another disadvantage associated with digital data is that researchers often lack control over the concepts and comprehensiveness of measurement. In the case of online behavior, monitoring a respondent's digital device through tracking cookies may not capture all of their online activity. For example, multiple cookies on user devices may be tracked at different times, which preserves some anonymity (Barthel et al., 2020). However, online behaviors are often repetitive and predictable over time, and recent studies with close proximity networks have shown that individuals may be identified even in pseudonymized datasets (Cretu et al., 2022). This example highlights the common tradeoff between comprehensiveness and privacy.

In the future, there will be increasing opportunities to repurpose alternative social datasets from the digital environment (Kalton, 2019; Rao 2021). Governments and their national statistical offices will be required to make important decisions regarding where and when to use conventional and digital data sources for policy. Survey data, though expensive, can be used to benchmark data from other sources that can be collected more cheaply, frequently, or with more granularity (Blumenstock et al., 2015; Keusch et al., 2020a, 2020b). The expansion of new sources in modern data collection including social data, sensors, and digital platforms are becoming serious complements and in some cases, alternatives, to conventional government surveys. These data sources are faster and cheaper to obtain through application programming interfaces but require increasingly complex tools to parse, handle, and compute. In the

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

following section, we describe four commonly employed technologies and their potential capabilities for near real-time analysis.

3. Technologies and analytics

The abundance of conventional and digital data allows for real-time monitoring and response while simultaneously posing data-processing challenges due to the vast amount of data available, also known as a "data avalanche". However, policymakers do not often face raw datasets, numeric and textual spreadsheets, or databases when crafting policy. They instead rely on the insights, knowledge, and evidence extracted from datasets through analytical processes.

The Technologies and Analytics Layer in Figure 2 encompasses existing and emerging technologies that link data collection and applications within policy. We do not explicitly differentiate between technology and analytics, as we find it arbitrary to draw a boundary between the two and note that it is not uncommon that an analytical module is underpinned by many different technologies. For example, Digital Twins and the IoT are often intertwined in the context of novel modeling, sensing, and data harvesting technologies, while from a system view, they sometimes also consist of analytical software components such as ML-enhanced IoT systems. Hence, the four widely used tools are blended under the umbrella of technologies and analytics in this section.

3.1. Machine learning

ML in CSS refers to the algorithms that allow computers to build predictions around behavioural data, thus "learning" and optimizing parameters over time. ML approaches may be supervised, in which the analyst provides labeled datasets to train the computer algorithm, or unsupervised, where the computer analyzes datasets without training on labeled data (Baraniuk et al., 2020). In recent years, there has been growth in the application of ML for policy problems, including supervised and unsupervised learning (Athey, 2017). Increasingly, governments are using supervised ML in prediction problems to determine how to best allocate resources. For instance, the New York City Fire Department's FireCast program uses ML to predict which buildings are most vulnerable to fire and deploy inspection teams (Heaton, 2015). Similar algorithms have been proposed for an increasing range of policy-relevant applications, such as environmental monitoring (Hino et al., 2018), preventing malfeasance in public procurement (Gallego et al., 2021), and restaurant hygiene inspections (Glaeser et al., 2016).

Researchers and analysts are also drawing on ML to leverage new datasets that capture hard-to-measure variables for policy analysis. Examples range from using Twitter to identify illegal sales of opioids (Mackey et al., 2017), developing economic uncertainty indices from scientific publications (Azqueta-Gavaldon, 2017), predicting income levels from phone metadata (Blumenstock et al., 2015), optimizing Covid-19 vaccine deployment strategies in Africa (Mellado et al., 2021), and predicting suicide risk using Reddit data (Yao et al., 2021; Allen et al., 2019). In many of these cases, researchers trained artificial intelligence (AI) to identify patterns (e.g., timing, wording, or events associated with behaviors of interest) from a small dataset and apply these algorithms to classify instances across a larger number of observations. Previously, tracking activities that are illegal or take place over large geographic areas would require a substantial and costly effort. Now, with deep neural networks, which are scalable and generalizable across domains, predictor variables from digital datasets can be observed or evaluated almost continuously (provided a robust, digitally available data feed). ML also enables researchers to utilize an abundance of unstructured data sources such as textual data, images, video, and other non-numeric data sources that might contain valuable information about the environment, policy preferences, or people's behaviors, among other aspects.

A major benefit of ML classification is the ability to link easily observed variables with more policy-relevant spatial or temporal qualities that may be harder to obtain. For example, in transportation infrastructure where there is poor data and network interoperability across jurisdictions, researchers have

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

been able to deploy deep learning algorithms to automatically detect failures in electric vehicle charging stations, with accuracy approaching or often exceeding that of human experts (Asensio et al., 2020; Ha et al., 2021). Deep learning algorithms have also been applied alongside satellite imagery to measure indicators of household consumption in poorer countries, where government statistics have more limited availability (Jean et al., 2016; Vinuesa and Sirmacek, 2021). More generally, deep neural networks such as transformer-based architectures (Vaswani et al., 2017; Yang et al., 2019; Devlin et al., 2018), and convolutional or recurrent neural networks (Gu et al., 2018; LeCun et al., 2015) are being deployed in an increasing number of policy-relevant applications to automatically discover context-aware, spatially resolved, and domain-specific insights (Hicks et al., 2022).

Large language models (e.g. GPT-4, BERT, BLOOM, Llama) have demonstrated strong performance in text generation. Since such technologies often emerge more rapidly than regulation or consensus around use policies can be determined, there is evidence of misuse (e.g., ChatGPT has been used to plagiarise on academic assignments and research studies and generate false narratives of misinformation (Else, 2023; Brewster et al., 2023)). To address this, several academic publishers, including Springer Nature and Cambridge University Press, have updated their policies to preclude generated text from attributed authorship (Thorpe, 2023; Cambridge University Press, CUP, 2023).

For policy domains, these AI systems are enhanced by humans who can intervene during training, testing, or validation to consider more complex issues (e.g. behavioral intent, psychological states, minority class representations, and other important social considerations). This blending of human and machine intelligence will also increase the need for high-quality, labeled training data that has been experimentally curated and approved by Institutional Review Boards (IRB) in sponsored research. There is a growing call for businesses to adopt IRB approval processes in data protocols as part of industry self-regulation and AI ethics boards (Blackman, 2021). We recognize that cross-disciplinary approaches that combine algorithmic advances with experimentally curated training data will continue to expand research frontiers and applications for ML/AI in social and policy domains.

Although there have been calls by the community for new regulations on applications of ML/AI, we note that there already exist mechanisms and applicable tools within the regulatory landscape, such as IRBs for the private sector and privacy regulations (e.g. General Data Protection Regulation (GDPR) in Europe and California Consumer Protection Act (CCPA) in the US), which can and have been leveraged to mitigate negative effects of ML/AI (Renieris, 2023). For example, the Information Commissioner's Office (ICO), which operates within the GDPR, prosecuted the facial recognition technology company Clearview AI for the unlawful collection and use of biometric data in 2022 (ICO, 2022). For further discussion, we refer readers to Focus Area 3: Policy and Literacy for Data and Focus Area 4: Ethics, Equity & Trustworthiness.

The growth of machine intelligence for policy decision-making also presents several challenges. First, social science inquiry that serves as a foundation for policy analysis typically focuses on developing a clear model explaining causal relationships. In contrast, many computational or purely data-driven approaches seek to optimize predictive performance, even if it makes the underlying model extremely complex or uninterpretable. Finding ways to integrate social theories with explainable AI systems (Amarasinghe et al., 2023) to illustrate why two variables are related at a micro level or better adapt interventions (Pallmann et al., 2018) is important to improve theoretically-driven, computational approaches (Hofman et al., 2021). Similarly, empirical constructs measured in real datasets often do not adequately capture the underlying social concepts they aim to articulate (Wagner et al., 2021), signaling a need for better validation and interpretation of measurements derived through ML (Buckee et al., 2021).

Second, many ML algorithms are seen as a "black box", wherein the actual mechanism linking data point A to prediction B is both hidden and not linked to a clear theoretical model in transfer functions. When working on policy questions that have real effects on health and well-being, these sorts of "trust the expert" approaches can erode trust in government and compliance (Coyle and Weller, 2020). This is especially true for supervised ML approaches, where the data scientist's

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

choice of algorithm and/or prediction weights suggests to the public that bias may be built into the prediction (Carmel, 2016).

Finally, while many in the *Data for Policy* community (policymakers and researchers alike) are excited about ML-powered data innovations, in many policy settings, the prediction enabled by ML may not be sufficient to fully answer existing policy questions. In some cases, this is because it is fundamentally a social question, wherein the computer can allocate a particular probability to a particular outcome, but society needs to decide what level of risk they are willing to accept (Kleinberg et al., 2015; Athey, 2017). In other cases, a stronger causal understanding is needed to develop effective policy solutions. However, integrating theories of policy and governance into ML policy analytics can be challenging because of the hands-off nature of the models and a historical over-reliance on nonexperimental data, where researchers do not precisely control the conditions during data collection, such as with random or quasi-random allocation mechanisms (Dunning, 2012). We are increasingly seeing effective uses of ML to target policy interventions, predict human behavior at a more granular level, and understand future events with greater precision (Amarasinghe et al., 2023). Despite practical issues on how best to apply these tools, there is an encouraging future for an increasing number of prediction policy problems.

3.2 Internet of Things

The IoT refers to systems in which computing and network-connected capabilities are incorporated within everyday objects to collect and distribute data without human intervention (Internet Society, 2015). The concept has evolved from a foundation in ubiquitous computing and includes various natural or built environment monitoring devices, wearable accessories, home appliances, vehicles, drones, smartphones, and computers (Satyanarayanan, 2001; Krumm, 2018; Friedewald and Raabe, 2011; Atzori et al., 2010). These technologies are capable of producing, transferring, and consuming real-time data about themselves and their surrounding environment. The collected data are then either streamed in real time or stored as historical data to be fed into ML and AI models for devising, monitoring, and evaluating various policies and regulations across different countries (Behrendt, 2020; Salem, 2017; Tanczer et al., 2019). On a more granular level, mobile sensing networks, which are comprised of numerous physical sensing nodes that record and wirelessly relay massive amounts of data, have been part of IoT systems for over two decades (Intanagonwiwat et al., 2000; Pottie and Kaiser, 2000; Tilak et al., 2002).

Cities around the world are increasingly responding to issues such as traffic congestion, air pollution, and natural hazards, which have prompted data-driven interventions to make cities more efficient, adaptable, and resilient. Over the past decades, the concept of IoT has been implemented in many real-world scenarios and 'smart' applications, particularly in the urban and city governance context, for example, smart transport and smart cities (Atzori et al., 2010). IoT-enabled decision or planning support systems are perhaps the most commonly used policy tools in the urban governance and smart city domain (Al Sharif and Pokharel, 2022). For instance, many cities have deployed various IoTs and sensor networks to cope with the natural hazards related to climate change (see Pantalona et al., 2021). Milton Keynes, UK is one of the cities that was selected by Innovate UK to test an IoT-empowered real-time air pollution monitoring system to support better public service for citizens and companies (Government Office for Science UK, 2014; Cheng et al., 2014). In India, IoT has been adopted in waste management for smart city scenarios (Sharma et al., 2020). IoT systems also act as a very crucial driving force in the global economy. For instance, bike-sharing systems (Behrendt, 2020) promote greener transport and support the UN Sustainable Development Goals (SDGs). Additionally, mobile sensing networks have been deployed since the 1990s to receive and transmit real-time data for pollution monitoring, satellite imaging and broadcasting, smart-home monitoring, and numerous other applications (Kahn et al., 1999; Dinh et al., 2013; Vermesan et al., 2022).

While bringing a number of social and economic benefits to governance and policy (Government Office for Science UK, 2014), IoT is still evolving and facing several major limitations. First, public concerns over data security and privacy protection have arisen from the existing applications of IoT (Tanczer et al., 2019; Ukil et al., 2014; Opara et al., 2022). Similar to many other big data technologies,

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

IoT systems collect, share, and analyze copious amounts of data about people and the environment, including sensitive information such as locations, trajectories, activities, health, and biometric data. The distributive architecture of IoT could also expose the sensors and devices to potential attacks and data intercepts, which undermine public trust in data security. In contrast with distributed IoT devices that may not have local control, personal mobile sensing devices could be more socially acceptable, considering their activities are more easily turned off (Choudhury et al., 2008). However, it is not transparent enough for people as data contributors to know what data are collected, where they are stored, and who has access to them (Corallo et al., 2022). Consequently, IoT devices present information problems related to user control and privacy.

Second, our capability of processing the data collected by various IoT systems lags behind the rate of data accumulation. As data is no longer a limitation in many applications, IoT-generated information about people and their environment is harvested continuously and is increasing in an exponential manner. The heterogeneous forms of data collected in texts, images, audio, and video materials further compound the complexity of data processing (Kazmi et al., 2018). Thus, it is still a challenging task to digest and refine the data and present more comprehensible knowledge to policymakers. There is a shortage of efficient and effective data analytical models as well as skilled IoT specialists, which is recognized in the governmental Blackett Review (Government Office for Science UK, 2014).

Third, there are a plethora of IoT solutions and applications for governance and regulation, while IoT systems themselves lack standards and policies. As a novel technology in multiple industries, IoT systems have many different domain-specific terminologies and standards. Countries may also have different focuses on their own views and definitions of IoT strategies. These inconsistencies have the potential to cause interoperability issues when coupling disparate IoT components on the internet, which is one of the barriers to adopting IoT more widely. Although a few countries have already explored and devised IoT-related regulations, for instance, India (Chatterjee and Kar, 2018), the UK (Tanczer et al., 2019), and the EU (Remotti et al., 2021), many regulations are still not IoT-specific and do not keep pace with the IoT evolution.

In response to the above challenges, future interactions between IoT and policymaking could be explored in several directions. In terms of data processing and knowledge discovery to inform policy-making processes, the IoT technology should be implemented with a more powerful and smarter analytical backend. For instance, utilizing ML and AI to extract meaningful patterns and evidence out of massive, heterogeneous datasets collected by IoT systems. This also requires a closer collaboration between policymakers and data experts (Government Office for Science UK, 2014). Policymakers should clearly present their real-world problems and define their queries to data experts, and in return, data experts need to inform them of both the findings of models and, more importantly, caveats that arise from possible data bias and model premises.

With respect to data security problems, IoT systems themselves should further embrace new digital technologies in data encryption and protection (Minoli and Occhiogrosso, 2018). Automatic security-enhancing mechanisms need to be implemented to detect essential or non-essential traffic over the IoT network, in order to restrict the transferring of sensitive information such as personally identifiable information without compromising the normal functionality of the IoT devices (Mandalari et al., 2021). Moreover, concerning regulation, policymakers should come up with more IoT-specific standards, policies, and laws (e.g., Government Office for Science UK, 2014) to guide the development of IoT. Policymaking and governance need to be more forward-thinking and adaptive to cope with the rapidly evolving IoT technologies.

3.3 Digital twins

The concept of Digital Twins, first coined by Michael Grieves at a Society of Manufacturing Engineers conference in 2003, has now proliferated beyond its origin in product lifecycle management into many other domains, including manufacturing, farming, healthcare, architecture, and city planning (Grieves, 2015). Unlike models and simulations, digital twins are more complex virtual environments that utilize

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

real-time data to generate multiple analyses (Bennett et al., 2023; Wright and Davidson, 2020). The Digital Twins ecosystem is underpinned by the various data sources mentioned in Section 2, as well as novel technologies such as sensor networks, IoT, 5G communication, cloud computing, ML, AI, virtual reality, augmented reality, mixed reality, geographic information systems (GISs), and building information modeling (BIM) (Wang et al., 2022). Recent years have seen the increased use of various digital twin scenarios and applications, particularly during the Covid-19 pandemic when people moved many physical, face-to-face activities to virtual cyberspace. There are several domain-specific definitions of Digital Twins, but in general, they are real-time, virtual representations of various physical or functional entities, and examples include digital human bodies, jet engines, buildings, infrastructures, and cities (Batty, 2018).

With innovative data and analytic techniques, the performance and dynamics of real-world entities can be measured, modeled, simulated, and predicted by their Digital Twins in virtual and software environments. These capabilities are effective and powerful tools for data-informed and evidence-based policymaking, particularly in the urban planning and management context (Engin et al., 2020). Digital Twins are employed to test various what-if scenarios for long-term urban planning and development. In order to achieve sustainable development, Digital Twins provide promising solutions to mitigate urban and regional issues such as poverty and inequalities (Birks et al., 2020), carbon footprints (Bauer et al., 2021; Solman et al., 2022), traffic congestion (Kumar et al., 2018), natural hazards (Fernández and Ceacero-Moreno, 2021), and public health problems (El Saddik et al., 2019). Digital twins are often difficult to replicate and scale. Recently, probabilistic graphical models have been proposed to ensure that digital twin representations and processes can be sufficiently scaled from experimental data to other physical assets (Mohammadi and Taylor, 2021; Kapteyn et al., 2021).

The applications of Digital Twins in governance and policymaking are numerous. During the Covid-19 pandemic, individual-level biometric data was collected and analyzed in the digital replica of the activity space to detect coexistence with the infected people (Ada Lovelace Institute, 2023). Smartphone-based digital contact tracing apps were developed and deployed to alert citizens of possible exposure in countries around the globe (Phillips et al., 2022). On a larger scale, Virtual Singapore is a government-led initiative, aiming to build a dynamic, three-dimensional, and city-scale Digital Twin of Singapore (Singapore Land Authority, n.d.). It enables different stakeholders, including members of the government, citizens, businesses, and the research community, to perform virtual experimentation, virtual test-bedding, long-term urban planning and decision-making, and research and development. Amaravati City in India is reported to be the first city that is newly developed on a greenfield site and born as a Digital Twin (Jansen, 2019). It ambitiously aims to digitally recreate everything happening in the city. For instance, it allows for real-time construction progress monitoring and advanced mobility simulations. The European Space Agency also launched several Digital Twin activities to visualize, monitor, model, and forecast natural and human activities, using earth observation data combined with AI, which would help human beings tackle pressing global issues such as climate change (European Space Agency, 2021).

Although Digital Twins offer promising platforms of data and policy interaction for integrating existing and emerging data sources and technologies, they also face many critiques and challenges, such as model difficulties (Tao and Qi, 2019) and scaling issues (Niederer et al., 2021). Scholars have also argued for the need to make more rapid adaptations in response to natural disasters and other challenges (Mohammadi and Taylor, 2021). Batty (2018) argues that most of the current computer models are abstractions or simplifications, rather than Digital Twins of the real world, and calls for a collaborative exploration as a society of how close our models can get to real-world systems. To address data privacy and availability problems, Papyshev and Yarime (2021) borrow the concept of "data labeling" from the ML industry practice and propose a task-based approach to generating synthetic data for City Digital Twins. On one hand, we call for the integration of novel data and technologies into Digital Twins to provide information and evidence for policymaking; on the other, we must pay more attention to the data and technology issues faced in Digital Twins implementation such as data infrastructure construction, data sharing, data security, privacy protection, interoperability, and platform standards, which can be regulated, directed and coordinated by relevant policies.

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

3.4 Blockchain and DLSs

Blockchain and DLSs are gaining attention in the government and business sectors due to their unique data-sharing features which are designed to increase transparency, authenticity, and reliability (Zutshi et al., 2021; Guo and Yu, 2022). DLSs refer to a digital framework that employs ledgers across multiple nodes or participants within a network, aiming to guarantee the security and accuracy of data (Marbouh et al., 2022). Blockchain emerged as an evolution of DLS, with an inherent capability to record transactions in chronological order in a secure and verifiable manner (Salah et al., 2019). Blockchain technologies append the data into "blocks" offering a range of benefits to support analytical applications, such as traceability, built-in anonymity, and secure transaction protocols (Mirabelli and Solina, 2020). Further, this technology offers decentralization that democratizes decision-making with no single authority in control (Beduschi, 2021). In particular, smart contract technology is gaining a growing focus due to its ability to streamline transaction processes and its potential for automating legal protocols (Hawashin et al., 2022). The features embedded in DLSs have the potential to bring change to the economic landscape with a new business model where the end customer is placed as the primary beneficiary (Upadhyay et al., 2021).

Fundamentally, two different types of blockchain networks have emerged, namely permissionless (i.e., public) and permissioned (i.e., private) networks (Engin and Treleaven, 2019). While any user can add nodes to the network in a public blockchain (e.g., Bitcoin), only preauthorized users can add nodes to a private blockchain network (e.g., Hyperledger Fabric) to reach consensus. For instance, many public networks use a Proof-of-Work (PoW) consensus mechanism with no single actor dominating it. However, public networks may suffer as more personal information will be required to verify the data added to the blockchain in attempts to prevent fraudulent activity. Further, although PoW ensures data immutability, its environmental and sustainability effects, such as bandwidth, electricity usage, and CPU time are significant challenges. In private networks, however, protocols are developed to better utilize computational resources. Despite such benefits, the challenge in private networks is to identify a technical solution to balance data verifiability and optimize the level of privacy among stakeholders. In both networks, another challenge is the increasing size of blockchains that may create storage and synchronization issues (Wong et al., 2021).

Various industries and domains, including finance (Tapscott and Tapscott, 2017), supply chain (Jabbar and Dani, 2020, Kayikci et al., 2022), and healthcare (Omar et al., 2020, Bali et al., 2022), explore the potential benefits of DLSs. Several academic initiatives aim to leverage the use of blockchain in various application areas. For instance, Cambridge Centre for Carbon Credits (4C) builds a trusted, decentralized voluntary carbon market for funding nature-based projects and seeks further partnerships with governments, the private sector, and NGOs to promote projects concerned with biodiversity and the climate crisis (Cambridge Zero Policy Forum, 2021). Despite the benefits of the technology, there have been a considerable number of failures in blockchain implementations. For instance, Browne (2017) shows that of the 26,000 blockchain projects that started in 2016, only 8% were still active in 2017. Various causes may explain the failure and hesitancy of the technology, mainly the hype around blockchain due to the volatility of cryptocurrencies (Jalal et al., 2021; Guo and Yu, 2022). The recent decline of cryptocurrencies and the downfall of major cryptoenterprises have raised further questions about the future of the technologies. In addition to these dramatic declines, concerns about money laundering, tax evasion attempts, and illicit payments have led financial services firms and venture capitalists to question the worth of investing in DLSs.

It should be noted that mainstream blockchain research primarily emphasizes technological aspects, often overlooking current regulatory functions. Although regulatory bodies have initiated working groups (e.g., the Australian Government National Blockchain Roadmap Working Group and European Blockchain Partnership), questions persist on the effectiveness and usability of legal mechanisms, particularly due to disintermediation (De Filippi et al., 2022). To address this challenge, various government organizations across the globe, such as Estonia (Ojo and Adebayo, 2017), the United Kingdom (Carson, 2018), the United Arab Emirates (Alketbi et al., 2020), and New Zealand (Demestichas et al.,

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

2020), started embracing the DLSs, particularly blockchain, as a strategic driver for technology and policy transformation. A recent study by IBM revealed that nine in ten government organizations explore opportunities to enhance their operations in different application areas, including financial transaction management, contract management, regulatory compliance, and citizen services (Cuomo et al., 2017).

Further application areas, such as in elections (Baudier et al., 2021) and vaccine passports to protect personal privacy (Tsoi et al., 2021), were also explored by governments to leverage the technology. To successfully develop and implement such applications, a suitable policy environment is imperative to support early collaborations between technology developers and policymakers and foster innovation compliance. For instance, some scholars support the idea that "minimum regulatory brakes" are the key to adding more value and efficiency to application areas (Yeoh, 2017). Such "hands-off" regulatory approaches have to date been adopted in the US and EU and show the potential for distributed trust frameworks. However, other scholars advocate for increased policy intervention, specifically on scalability, privacy, security, sustainability, and anonymity (Hassan et al., 2020; Liiv, 2021). Considering the potential benefits and challenges of the technology, policy environments may enable experimentation (McQuinn and Castro, 2019) and learn from the experiences of others in the global landscape (PWC, 2019) to recommend informed regulatory changes accordingly. Future studies may benefit from exploring the potential impact of DLSs and blockchain in the entire technology sector and disruption in government and business operations and policymaking.

4. Using new data sources and analytics in policy-making

Evidence-based policy refers to the efforts to prioritize data-based decision-making in policy processes (Head, 2008; Howlett, 2009; Evidence-Based Policymaking Collaborative, 2016). The proliferating data sources and analytical techniques made available through the big data science enable new ways of bringing evidence to the design, implementation, and monitoring of policies and programs (Anshari et al., 2018; Kim et al., 2014; Giest, 2017; Suominen and Hajikhani, 2021). However, it is important to note that these advances do not necessarily translate into automatic uptake by policymakers. Governments face numerous constraints, from limited budgets, to external political and social pressures, to varying technical expertise, that limit their capability to fully capitalize upon the information available (Mergel et al., 2016; Schweinfest and Jansen, 2021). This is true not only of new sources of data and innovative methodologies, but as a more general concern that also traverses traditional evidence-based approaches. How governments draw on data to inform decision-making is covered in greater depth by Data for Policy Focus Area 1: Digital and Data-Driven Transformations in Governance.

The data and analytical approaches discussed in sections 2 and 3 raise several unique challenges with respect to their uptake in evidence-based decision-making. First, governments tend to lag behind the private sector in adopting new computing technologies (Dunleavy et al., 2006). Given the rapid pace of advancements in data and analytics, government agencies are often delayed in adopting the newest approaches. Second, many government workers perceive a skills gap in the use of data and analytics, despite viewing data as a central component of their jobs (SAS, 2014). The World Economic Forum Jobs Report estimates that 24 percent of government and public sector organizations are making big data and AI a reskilling priority (World Economic Forum, WEF, 2023). When government agencies lack data expertise but want to use new analytical approaches, they have to rely on other actors (primarily from the private sector), leading to the growth of public–private partnerships for data (Geist, 2017). These public-private partnerships add to institutional complexity (Head, 2008) but also offer opportunities for innovation (Janssen et al., 2017). Third, some government administrators lack an understanding of what data analytics entails or are skeptical about its ability to address policy problems (Guenduez et al., 2020). Fourth, the integration of big data can vary widely between developed and developing countries due to challenges in basic data availability and skills within the public sector. Applications are more common in developed countries, where access to data technology skills is more readily available (Purkayastha and Braa, 2013). Additional known challenges that could hinder developing countries from integrating the

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

same technologies include limited data capture, infrastructural constraints, human resource scarcity, privacy and security constraints, and cultural barriers (Luna et al., 2014; Hilbert, 2015).

Despite these limitations, there are many examples of government agencies drawing on new data sources and analytical techniques, as we have illustrated throughout this paper. New training programs are helping government employees build expertise in data analytics which could overcome existing skills gaps (Kreuter et al., 2019). Further uptake of these approaches could be facilitated by policymakers, analytical experts, and members of policy-affected communities collaboratively identifying data needs and codeveloping analytical approaches, as the coproduction of knowledge is known to result in credible, salient, and trusted information (Ulibarri, 2018; Cravens, 2016; Morisette et al., 2017).

5. Closing

The *Data for Policy* community is contributing to innovations in digital data use and supporting technologies and analytics for policy decision-making. We conclude this initial landscape report with three observations highlighted by our community: there is a need for a greater emphasis on (i) model explainability, (ii) broader cross-sector collaboration, and (iii) data accessibility.

First, we note that without the integration of appropriate social science theories or hypothesis testing to guide feature selection in computational modeling, there is often the "black box" temptation to model phenomena using fully data-driven approaches. Although this continues to be very useful in domains (e.g. cancer detection, pollution monitoring, etc.), a greater focus on causal inference can help prevent the social ills sometimes observed in algorithmic decision-making (for additional information, see Veale et al., 2018; Data for Policy Focus Area 5: Algorithmic Governance).

Our second observation is the need to increase cross-sector collaborations. Broadening this network between academics and practitioners is especially important as significant decisions regarding the use of personal data are being made largely outside of academia. In addition, such collaborations could benefit from more direct engagement with representatives of the affected communities (both positively and negatively) and/or the general public as a way to increase trust and reduce unintended side effects. These new models of scientific collaboration will be beneficial to catalyzing principled engagement for data-informed decision-making within the public sector.

The third observation relates to challenges associated with data accessibility and preserving anonymity. A majority of digital data sources are concentrated in platforms controlled by private companies, often inaccessible to government agencies. Consequently, there continue to be significant legal and financial barriers to accessing this data even when there is a compelling need (Salganik, 2019). In addition, digital data presents significant challenges related to incidental disclosure or re-identification and we have recently learned that personally identifiable information can be recovered even from anonymized or pseudonymized datasets (Kearns and Roth, 2019; De Montjoye et al., 2015; Cretu et al., 2022). Numerous approaches have been developed to address this problem, most notably the framework of differential privacy, but its application to standard social datasets has been met with criticism and limits related to data integrity (Cummings et al., 2019; Dwork, 2008; Abowd, 2018; Dwork, 2019; Ruggles et al., 2019). These issues are addressed in further detail in Data for Policy Focus Area 4: Ethics, Equity, & Trustworthiness.

The Area 2 committee will be focusing on manuscripts that investigate the impacts of LLMs and other generative AI models and how regulators can respond to ensure sound decision-making that benefits humanity and societies. We invite authors working at the data science-policy interface to engage with the community through the Data for Policy conference series and submissions to the Data and Policy journal.

Acknowledgments. We thank members of the editorial board and the *Data for Policy* community who provided valuable discussion and reviews.

Author contribution. Conceptualization: O.I.A., T.L., C.M., G.R., M.C.E.S., N.U.; Methodology: O.I.A., T.L., C.M., G.R., M.C.E.S., N.U.; Data curation: O.I.A., M.C.E.S.; Data visualization: O.I.A., T.L., C.M.; Investigation: O.I.A., T.L., C.M., G.R., M.C.E.S., N.U.; Project administration: O.I.A., C.M.; Validation: O.I.A., C.M., N.U.; Writing—original draft: O.I.A., T.L., C.M., G.R., M.C.E.S., N.U.; Writing—review & editing: O.I.A., C.M., N.U. All authors approved the final submitted draft.

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

Data availability statement. Manuscript submission data that support the findings of this study are available upon request from dataandpolicy@cambridge.org. The data are not publicly available due to author privacy restrictions for double-blind peer review.

Provenance. This article was authored by the Editors associated with Data for Policy Focus Area 2: Data Technologies & Analytics. It was independently reviewed.

Funding statement. O.I.A. and C.M. were partially supported by National Science Foundation Award #1945332. The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interest. The authors declare no competing interests.

References

Abowd JM (2018) The U.S. Census Bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*. New York, NY, USA: Association for Computing Machinery, 2867. https://doi.org/10.1145/3219819.3226070

Abul-Husn NS and Kenny EE (2019) Personalized medicine and the power of electronic health records. *Cell 177*(1), 58–69. https://doi.org/10.1016/j.cell.2019.02.039.

Ada Lovelace Institute (ALI) (2023) Lessons from the App Store: Insights and learnings from COVID-19 technologies Report. Available at https://www.adalovelaceinstitute.org/wp-content/uploads/2023/06/Ada-Lovelace-Institute-Lessons-from-the-App-Store-June-2023.pdf (accessed 10 August 2023).

Aguilera A and Boutueil V (2018) *Urban Mobility and the Smartphone: Transportation, Travel Behavior and Public Policy.* Elsevier.

Albino V, **Umberto B and Dangelico RM** (2015) Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology 22*(1), 3–21. http://doi.org/10.1080/10630732.2014.942092.

Allen K, **Bagroy S**, **Davis A and Krishnamurti T** (2019) ConvSent at CLPsych 2019 task a: using post-level sentiment features for suicide risk prediction on reddit. *In Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology.* Minneapolis: Association for Computational Linguistics, 182–187. https://doi.org/10.18653/v1/W19-3024.

Alketbi A, **Nasir Q and Abu Talib M** (2020) Novel blockchain reference model for government services: Dubai government case study. *International Journal of System Assurance Engineering and Management 11*, 1170–1191. https://doi.org/10.1007/s13198-020-00971-2

Al Sharif R and Pokharel S (2022) Smart city dimensions and associated risks: Review of literature. *Sustainable Cities and Society 77*.https://doi.org/10.1016/j.scs.2021.103542

Amarasinghe K, **Rodolfa K**, **Lamba H and Ghani R** (2023). Explainable machine learning for public policy: Use cases, gaps, and research directions. *Data & Policy 5*, E5. https://doi.org/10.1017/dap.2023.2

Angrist JD and Pischke JS (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30. https://doi.org/10.1257/jep.24.2.3

Anshari M, **Almunawar MN and Lim SA** (2018). Big data and open government data in public services. In *Proceedings of the 2018 10th International Conference on Machine Learning and Computing.* New York: Association for Computating Machinery, 140–144. https://doi.org/10.1145/3195106.3195172

Arora VS, **McKee M and Stuckler D** (2019) Google Trends: Opportunities and limitations in health and health policy research. *Health Policy 123*(3), 338–341. https://doi.org/10.3389/fdata.2023.1132764

Asensio OI, **Alvarez K**, **Dror A**, **Wenzel E**, **Hollauer C and Ha S** (2020) Real-time data from mobile platforms to evaluate sustainable transportation infrastructure. *Nature Sustainability 3*, 463–471. https://doi.org/10.1038/s41893-020-0533-6

Asensio OI, **Apablaza CZ**, **Lawson MC**, **Chen EW and SJ Horner** (2021) Impacts of micromobility on car displacement with evidence from a natural experiment and geofencing policy. *Nat Energy 7*, 1100–1108. https://doi.org/10.1038/s41560-022-01135-1

Athey S and Imbens G (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences 113*(27), 7353–7360. https://doi.org/10.1073/pnas.151048911

Athey S and Imbens G (2017) The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, *31*(2), 3–32. https://doi.org/10.1257/jep.31.2.3

Athey S (2017) Beyond prediction: using big data for policy problems. *Science 355*(6324), 483–485. https://doi.org/10.1126/science.aal4321

Atzori L, **Iera A and Morabito G** (2010) The internet of things: A survey. *Computer Networks*, *54*(15), 2787–2805. https://doi.org/10.1016/j.comnet.2010.05.010

Azqueta-Gavaldon A (2017) Developing news-based economic policy uncertainty index with unsupervised machine learning. *Economics Letters 158*, 47–50. https://doi.org/10.1016/j.econlet.2017.06.032

Bali B, **Bali V**, **Mohanty R and Gaur D** (2022) Analysis of critical success factors for blockchain technology implementation in healthcare sector. *Benchmarking: An International Journal 30*(1). https://doi.org/10.1108/BIJ-07-2021-0433

Baraniuk R, **Donoho D and Gavish M** (2020) The science of deep learning. *Proceedings of the National Academy of Sciences 117* (48), 30029–30032. https://doi.org/10.1073/pnas.2020596117

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

Barberá P and Rivero G (2014) Understanding the political representativeness of Twitter users. *Social Science Computer Review 33*(6), 712–729. https://doi.org/10.1177/0894439314558836

Barthel M, **Mitchell A**, **Asare-Marfo D**, **Kennedy C and Worden K** (2020) Measuring news consumption in a digital era. *Pew Research Center's Journalism Project*, *8* Available at https://www.pewresearch.org/journalism/2020/12/08/measuring-news-consumption-in-a-digital-era/

Batty M (2018). Digital twins. *Environment and Planning B: Urban Analytics and City Science 45*(5), 817–820. https://doi.org/10.1177/2399808318796416

Batty M, **Axhausen KW**, **Giannotti F**, **Pozdnoukhov A**, **Bazzani A**, **Wachowicz M**, **Ouzounis G and Portugali Y** (2012) Smart cities of the future. *The European Physical Journal Special Topics 214*(1), 481–518. https://doi.org/10.1140/epjst/e2012-01703-3

Baudier P, **Kondrateva G**, **Ammi C and Seulliet E** (2021) Peace engineering: The contribution of blockchain systems to the e-voting process. *Technological Forecasting and Social Change 162* https://doi.org/10.1016/j.techfore.2020.120397

Bauer P, **Stevens B and Hazeleger W** (2021) A digital twin of Earth for the green transition. *Nature Climate Change 11*(2), 80–83. https://doi.org/10.1038/s41558-021-00986-y

Beduschi A (2021) Rethinking digital identity for post-COVID-19 societies: Data privacy and human rights considerations. *Data & Policy 3*, E15. https://doi.org/10.1017/dap.2021.15

Behrendt F (2020) Mobility and data: Cycling the utopian Internet of Things. *Mobilities 15*(1), 81–105. https://doi.org/10.1080/17450101.2019.1698763

Bennett H, **Birkin M**, **Ding J**, **Duncan A and Engin Z** (2023) *Towards Ecosystems of Connected Digital Twins to Address Global Challenges*. Alan Turing Institute. https://doi.org/10.5281/zenodo.7840266

Birks D, **Heppenstall A and Malleson N** (2020) Towards the development of societal twins. In *Frontiers in Artificial Intelligence and Applications.* Santiago de Compostela: 24th European Conference on Artificial Intelligence, 2883–2884. https://doi.org/10.3233/FAIA200435

Blackman R (2021) *If Your Company Uses AI, It Needs an Institutional Review Board.* Harvard Business Review, Business Ethics. Available at https://hbr.org/2021/04/if-your-company-uses-ai-it-needs-an-institutional-review-board (accessed 30 August 2023).

Blair G and Imai K (2012) Statistical analysis of list experiments. *Political Analysis 20*(1), 47–77. https://doi.org/10.1093/pan/mpr048

Blumenstock J, **Cadamuro G and On R** (2015) Predicting poverty and wealth from mobile phone metadata. *Science 350*(6264), 1073–76. https://doi.org/10.1126/science.aac4420

Brewster J, **Arvanitis L and Sadeghi M** (2023) Misinformation Monitor: January 2023. *Newsweek*, 15 February. Available at https://www.newsweek.com/2023/02/24/misinformation-monitor-january-2023-1781533.html (accessed 2 March 2023).

Brick JM and Williams D (2013) Explaining rising nonresponse rates in cross-sectional surveys. *The ANNALS of the American Academy of Political and Social Science 645*(1), 36–59. https://doi.org/10.1177/0002716212456834

Browne R (2017) There were more than 26,000 new blockchain projects last year – only 8% are still active. *CNBC*, 9 November. Available at https://www.cnbc.com/2017/11/09/just-8-percent-of-open-source-blockchain-projects-are-still-active.html (accessed 8 August 2021).

Buckee C, **Noor A and Sattenspiel L** (2021) Thinking Clearly about Social Aspects of Infectious Disease Transmission. *Nature 595*(7866), 205–213. https://doi.org/10.1038/s41586-021-03694-x

Cambridge University Press (2023) *Authorship and Contributorship.* Available at https://www.cambridge.org/core/services/authors/publishing-ethics/research-publishing-ethics-guidelines-for-journals/authorship-and-contributorship#ai-contributions-to-research-content (accessed 29 March 2023).

Cambridge Zero Policy Forum (2021) *Carbon Offsetting & Nature-based Solutions to Climate Change.* University of Cambridge. Available at https://www.csap.cam.ac.uk/media/uploads/files/1/cambridge-zero-policy-forum-discussion-paper-carbon-offsetting-and-nature-based-solutions-to-climate-change.pdf

Carmel YH (2016) Regulating "big data education" in Europe: Lessons learned from the US. *Internet Policy Review 5*(1).https://doi.org/10.14763/2016.1.402

Carson K (2018) Blockchain will open up the global retail banking system. *London School of Economics Business Review*, *2*. Available at https://blogs.lse.ac.uk/businessreview/2018/11/02/blockchain-will-open-up-the-global-retail-banking-system/ (accessed 10 August 2021).

Cebul RD, **Love TE**, **Jain AK and Hebert CJ** (2011) Electronic health records and quality of diabetes care. *New England Journal of Medicine 365*(9), 825–833. https://doi.org/10.1056/nejmsa1102519

Chatterjee S and Kar AK (2018) Regulation and governance of the Internet of Things in India. *Digital Policy, Regulation and Governance 20*. https://doi.org/10.1108/DPRG-04-2018-0017

Cheng Y, **Li X**, **Li Z**, **Jiang S**, **Li Y**, **Jia J and Jiang X** (2014) AirCloud: A cloud-based air-quality monitoring system for everyone. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, 251–265. https://doi.org/10.1145/2668332.2668346

Choudhury T, **Borriello G**, **Consolvo S**, **Haehnel D**, **Harrison B**, **Hemingway B**, **Hightower J**, **Klasnja PP**, **Koscher K**, **LaMarca A**, **Landay JA**, **LeGrand L**, **Lester J**, **Rahimi A**, **Rea A and Wyatt D** (2008) The Mobile Sensing Platform: An Embedded Activity Recognition System. *IEEE Pervasive Computing 7*, 32–41. https://doi.org/10.1109/MPRV.2008.39

Connelly R, **Playford CJ**, **Gayle V and Dibben C** (2016) the role of administrative data in the big data revolution in social science research. *Social Science Research 59*, 1–12. https://doi.org/10.1016/j.ssresearch.2016.04.015

Corallo A, **Lassi M**, **Lezzi M and Luperto A** (2022) Cybersecurity awareness in the context of the Industrial Internet of Things: A systematic literature review. *Computers in Industry 137*. https://doi.org/10.1016/j.compind.2022.103614

Cowie MR, **Blomster JI**, **Curtis LH**, **Duclaux S**, **Ford I**, **Fritz F and Goldman S** (2017) Electronic Health Records to Facilitate Clinical Research. *Clinical Research in Cardiology 106*(1), 1–9. https://doi.org/10.1007/s00392-016-1025-6

Coyle D and Weller A (2020) Explaining machine learning reveals policy challenges. *Science 368*(6498), 1433–1434. https://doi.org/10.1126/science.aba9647

Cravens AE (2016) Negotiation and decision making with collaborative software: How MarineMap 'changed the game'in California's Marine Life Protected Act Initiative. *Environmental Management*, *57*(2), 474–497. https://doi.org/10.1007/s00267-015-0615-9

Crețu AM, **Frederico M**, **Marrone S**, **Dong X**, **Bronstein M and De Montjoye Y** (2022) Interaction data are identifiable even across long periods of time. *Nature Communications 13*(313). https://doi.org/10.1038/s41467-021-27714-6

Cummings R, **Gupta V**, **Kimpara D and Morgenstern, J** (2019) *On the Compatibility of Privacy and Fairness.* pp.309–315. https://doi.org/10.1145/3314183.3323847

Cuomo J, **Pureswaran V and Zaharchuk D** (2017) Building trust in government: Exploring the potential of blockchains. Available at https://www.ibm.com/thought-leadership/institute-business-value/report/blockchain-for-government

Demestichas K, **Peppes N**, **Alexakis T and Adamopoulou E** (2020) Blockchain in agriculture traceability systems: A review. *Applied Sciences 10*(12). https://doi.org/10.3390/app10124113

Devlin J, **Chang M**, **Lee K and Toutanova K** (2018) BERT: Pre-training of deep bidirectional transformers for language understanding. *ArXiv* preprint 1810.04805v2. Available at https://arxiv.org/pdf/1810.04805.pdf

De Filippi P, **Mannan M and Reijers W** (2022) The alegality of blockchain technology. *Policy and Society 41*(3) 358–372, https://doi.org/10.1093/polsoc/puac006

De Montjoye Y, **Redaelli L**, **Kumar Singh V and Pentland A** (2015) Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science 347*(6221), 536–539. https://www.science.org/doi/10.1126/science.1256297

DiGrazia J, **McKelvey K**, **Bollen J and Rojas F** (2013) More tweets, more votes: Social media as a quantitative indicator of political behavior. *PloS One 8*(11). http://doi.org/10.1371/journal.pone.0079449

Dinh HT, **Lee C**, **Niyato D and Wang P** (2013) A survey of mobile cloud computing: architecture, applications, and approaches. *Wireless Communications and Mobile Computing*, *13*(18), 1587–1611.

Dong L and Shan J (2013) A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing 84*, 85–99. https://doi.org/10.1016/j.isprsjprs.2013.06.011

Dugas AF, **Jalalpour M**, **Gel Y**, **Levin S**, **Torcaso F**, **Igusa T and Rothman RE** (2013) Influenza forecasting with google flu trends. *PloS One 8*(2). https://doi.org/10.1371/journal.pone.0056176

Dunleavy P, **Margetts H**, **Bastow S and Tinkler J** (2006) Digital era governance: IT corporations, the state, and E-Government. *Social Science Computer Review 26*(2), 254–257. https://doi.org/10.1177/0894439307304515

Dunning T (2012) *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge University Press. https://doi.org/10.1017/CBO9781139084444

Dwork C (2008) Differential privacy: A survey of results. In International Conference on Theory and Applications of Models of Computation, Xi'an, China, 25-29 April 2008, 1–19. https://doi.org/10.1007/978-3-540-79228-4_1

Dwork C, **Kohli N and Mulligan D** (2019). Differential privacy in practice: Expose your epsilons!. *Journal of Privacy and Confidentiality 9*(2). https://doi.org/10.29012/jpc.689

U.S. Energy Information Administration (EIA) (2022) Residential Energy Consumption Survey (RECS). Available at https://www.eia.gov/consumption/residential/

Engin Z, **Gardner E**, **Hyde A**, **Verhulst SV and Crowcroft J** (2024) unleashing collective intelligence for public decision making: The data for policy community. *Data & Policy 6*:e2. https://doi.org/10.1017/dap.2024.2

Engin Z and Treleaven P (2019) Algorithmic Government: Automating public services and supporting civil servants in using data science technologies. *The Computer Journal 62*(3), 448–460. https://doi.org/10.1093/comjnl/bxy082

Engin Z, **van Dijk J**, **Lan T**, **Longley PA**, **Treleaven P**, **Batty M and Penn A** (2020) Data-driven urban management: Mapping the landscape. *Journal of Urban Management 9*(2), 140–150. https://doi.org/10.1016/j.jum.2019.12.001

El Saddik A, **Badawi H**, **Velazquez RAM**, **Laamarti F**, **Diaz RG**, **Bagaria N and Arteaga-Falconi JS** (2019) Dtwins: a digital twins ecosystem for health and well-being. *Artificial Intelligence Methodologies for Networked Sensors in Smart Cities 14*, 39–43. https://doi.org/10.3390/s21041047

Else H (2023) Abstracts written by ChatGPT fool scientists. *Nature.* Available at https://www.nature.com/articles/d41586-023-00056-7 (accessed 2 March 2023).

Elliott MR and Valliant R (2017) Inference for nonprobability samples. *Statistical Science 32*(2), 249–264. https://doi.org/10.1214/16-STS598

European Space Agency (2021) *Working Towards a Digital Twin of Eart*h. Available at: https://www.esa.int/Applications/Observing_the_Earth/Working_towards_a_Digital_Twin_of_Earth (accessed 22 February 2022)

Evidence-Based Policymaking Collaborative (2016) *Principles of Evidence-Based Policymaking, September 2016.* Available at https://www.urban.org/sites/default/files/publication/99739/principles_of_evidence-based_policymaking.pdf

Ferguson AG (2017) *The Rise of Big Data Policing*. New York University Press.

Fernández P and Ceacero-Moreno M (2021) Urban sustainability and natural hazards management; designs using simulations. *Sustainability 13*(2), 649. https://doi.org/10.3390/su13020649

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

Friedewald M and Raabe O (2011). Ubiquitous computing: An overview of technology impacts. *Telematics and Informatics 28* (2), 55–65. https://doi.org/10.1016/j.tele.2010.09.001

Gallego J, **Rivero G and Martínez J** (2021) Preventing rather than Punishing: An early warning model of malfeasance in public procurement. *International Journal of Forecasting 37*(1), 360–377. https://doi.org/10.1016/j.ijforecast.2020.06.006

Gandy M, **Baker PMA and Zeagler C** (2017) Imagining Futures: A Collaborative Policy/Device Design for Wearable Computing. *Futures 87*(5), 106–121. https://doi.org/10.1016/j.futures.2016.11.004

Gayo-Avello D, **Metaxas PT and Mustafaraj E** (2011) Limits of electoral predictions using twitter. In *Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona.

Giest S (2017) Big data for policymaking: Fad or fasttrack? *Policy Sciences 50*(3), 367–382. https://doi.org/10.1007/s11077-017-9293-1

Glaeser EL, **Hillis A**, **Kominers SD and Luca M** (2016) Crowdsourcing city government: Using tournaments to improve inspection accuracy. *The American Economic Review 106*(5), 114–118. https://doi.org/10.1257/aer.p20161027

Golder SA and Macy MW (2014) Digital footprints: opportunities and challenges for online social research. *Annual Review of Sociology 40*,129–152. http://doi.org/10.1146/annurev-soc-071913-043145

Government Office for Science UK (2014). Internet of things: Blackett review. Available at: https://www.gov.uk/government/publications/internet-of-things-blackett-review (accessed 18 November 2021)

Green D, **Moszczynski M**, **Asbah S**, **Morgan C**, **Klyn B**, **Foutry G**, **Ndira S**, **Selman N**, **Monawe M**, **Likaka A**, **Sibande R and Smith T** (2021) Using mobile phone data for epidemic response in low resource settings—a Case Study of COVID-19 in Malawi. *Data & Policy 3*, E19. https://doi.org/10.1017/dap.2021.14

Grieves M (2015) Digital Twin: Manufacturing excellence through virtual factory replication. *A whitepaper.* Available at: https://www.researchgate.net/publication/275211047_Digital_Twin_Manufacturing_Excellence_through_Virtual_Factory_Replication (accessed 23 July 2021).

Grotpeter J (2008) Respondent recall. In *Handbook of Longitudinal Research: Design, Measurement, and Analysis.* Academic Press, 109–122.

Groves RM (2011) Three Eras of Survey Research. *Public Opinion Quarterly 75*(5), 861–171. https://doi.org/10.1093/poq/nfr057

Gu J, **Wang Z**, **Kuen J**, **Ma L**, **Shahroudy A**, **Shuai B**, **Liu T**, **Wang X**, **Wang G**, **Cai J and Chen T** (2018) Recent advances in convolutional neural networks. *Pattern Recognition 77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013.

Guenduez AA, **Mettler T and Schedler K** (2020) Technological frames in public administration: What do public managers think of big data? *Government Information Quarterly 37*(1), 101406. https://doi.org/10.1016/j.giq.2019.101406

Guo H and Yu X (2022) A survey on blockchain technology and its security. *Blockchain: Research and Applications 3* (2).https://doi.org/10.1016/j.bcra.2022.100067

Ha S, **Marchetto DJ**, **Dharur S and Asensio OI** (2021) Topic classification of electric vehicle consumer experiences with transformer-based deep learning. *Patterns 2*, 100195. https://doi.org/10.1016/j.patter.2020.100195

Hassan F, **Ali A**, **Rahouti M**, **Latif S**, **Kanhere S**, **Singh J**, **Janjua U**, **Mian AN**, **Qadir J and Crowcroft J** (2020) Blockchain and the future of the internet: A comprehensive review. arXiv preprint 1904.00733. Available at https://arxiv.org/abs/1904.00733

Hawashin, D; **Jayaraman, R**; **Salah, K**; **Yaqoob, I**; **Simsekler, M. C. E**; **Ellahham, S** (2022) Blockchain-based management for organ donation and transplantation. *IEEE Access 10*, 59013–59025. https://doi.org/10.1109/ACCESS.2022.3180008.

Head B (2008) Three lenses of evidence-based policy. *Australian Journal of Public Administration 67*(1), 1–11. https://doi.org/10.1111/j.1467-8500.2007.00564.x

Heaton B (2015) New York City Fights Fire with Data. GovTech article. Available at https://www.govtech.com/public-safety/new-york-city-fights-fire-with-data.html

Hicks D, **Zullo M**, **Doshi A and Asensio OI** (2022) Widespread use of National Academies consensus reports by the American public. *Proceedings of the National Academy of Sciences 119*(9), e2107760119, https://doi.org/10.1073/pnas.2107760119

Hilbert M (2015) Big data for development: A review of promises and challenges. *Development Policy Review 34*(1), 135–174. https://doi.org/10.1111/dpr.12142

Hino M, **Benami E and Brooks N** (2018) Machine learning for environmental monitoring. *Nature Sustainability 1*(10) 583–588. https://doi.org/10.1038/s41893-018-0142-9

Hofman JM, **Watts DJ**, **Athey S**, **Garip F**, **Griffiths TL**, **Kleinberg J**, **Margetts H**, **Mullainathan S**, **Salganik MJ**, **Vazire S**, **Vespignani A and Yarkoni T** (2021) Integrating explanation and prediction in computational social science. *Nature* **595**(7866), 181–188. https://doi.org/10.1038/s41586-021-03659-0

Howlett M (2009) Policy analytical capacity and evidence-based policy-making: Lessons from Canada. *Canadian Public Administration 52*(2), 153–175. https://doi.org/10.1111/j.1754-7121.2009.00070_1.x

Information Commissioner's Office (ICO) (2022) ICO fines facial recognition database company Clearview AI Inc more than £7.5m and orders UK data to be deleted. Available at https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2022/05/ico-fines-facial-recognition-database-company-clearview-ai-inc/ (accessed 10 August 2023).

Intanagonwiwat C, **Govindan R and Estrin D** (2000, August) Directed diffusion: A scalable and robust communication paradigm for sensor networks. In *Proceedings of the 6th annual international conference on Mobile computing and networking* (pp. 56–67). Boston, Massachusetts, USA: Association for Computing Machinery.

Internet Society (2015) *The Internet of Things: An Overview. Understanding the Issues and Challenges of a More Connected World.* Available at https://www.internetsociety.org/wp-content/uploads/2017/08/ISOC-IoT-Overview-20151221-en.pdf. (accessed 10 August 2023).

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

Jabbar A and Dani S (2020) Investigating the link between transaction and computational costs in a blockchain environment. *International Journal of Production Research 58*, 3423–3436. https://doi.org/10.1080/00207543.2020.1754487

Janssen M, **Konopnicki D**, **Snowdon J and Ojo A** (2017) Driving public sector innovation using big and open linked data (BOLD). *Information Systems Frontiers 19*(2), 189–195. https://doi.org/10.1007/s10796-017-9746-2

Jansen M (2019) *Digital Twins for Greenfield Smart Cities.* Available at https://newcities.org/the-big-picture-digital-twins-for-greenfield-smart-cities/ (accessed 24 July 2021).

Jean N, **Burke M**, **Xie M**, **Davis WM**, **Lobell DB, and Ermon S** (2016) Combining Satellite Imagery and Machine Learning to Predict Poverty. *Science 353*(6301), 790–94. https://doi.org/10.1126/science.aaf7894

Jungherr A, **Rivero G and Gayo-Avello D** (2020) Retooling politics: How digital media are shaping democracy. *Cambridge University Press* **27**(2). https://doi.org/10.1177/19401612221073994

Kahn JM, **Katz RH and Pister KS** (1999) Next century challenges: mobile networking for "Smart Dust". In *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking* (pp. 271–278).

Kalton G (2019) Developments in Survey Research over the Past 60 Years: A Personal Perspective. *International Statistical Review 87*, S10–S30. https://doi.org/10.1111/insr.12287

Kapteyn MG, **Pretorius JVR and Willcox KEA** (2021) Probabilistic graphical model foundation for enabling predictive digital twins at scale. *Nature Computational Science 1*, 337–347. https://doi.org/10.1038/s43588-021-00069-0

Kayikci Y, **Gozacan-Chase N**, **Rejeb A and Mathiyazhagan K** (2022) Critical success factors for implementing blockchain-based circular supply chain. *Business Strategy and the Environment 31*(7), 3595–3615. https://doi.org/10.1002/bse.3110

Kazmi A, **Serrano M and Lenis A** (2018) Smart governance of heterogeneous internet of things for smart cities. In *12th International Conference on Sensing Technology.*

Kearns M and Roth A (2019) *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford University Press.

Kennedy C, **Mercer A**, **Keeter S**, **Hatley N**, **McGeeney K and Gimenez A** (2016) Evaluating online nonprobability surveys. *Pew Research.* Available at https://www.pewresearch.org/methods/2016/05/02/evaluating-online-nonprobability-surveys/

Keusch F, **Struminskaya B**, **Kreuter F and Weichbold M** (2020a) *Combining Active and Passive Mobile Data Collection: A Survey of Concerns.* Big Data Meets Survey Science: A Collection of Innovative Methods.

Keusch F, **Bähr S**, **Haas G**, **Kreuter F and Trappmann M** (2020b) Coverage error in data collection combining mobile surveys with passive measurement using apps: Data from a German National survey. *Sociological Methods & Research 52*(17). https://doi.org/10.1177/0049124120914924

Khalaf M, **Abir JH**, **Al-Jumeily D**, **Fergus P and Idowu IO** (2015) Advance flood detection and notification system based on sensor technology and machine learning algorithm. In 2015 *International Conference on Systems, Signals and Image Processing (IWSSIP).* pp. 105–108.

Kim G, **Trimi S and Chung J** (2014) Big-data applications in the government sector. *Communications of the ACM 57*(3), 78–85. https://doi.org/10.1145/2500873

Klašnja M, **Barberá P**, **Beauchamp N**, **Nagler J and Tucker JA** (2015) Measuring public opinion with social media data. In *The Oxford Handbook of Polling and Polling Methods*,. Oxford University Press,, 555–582

Kleinberg J, **Ludwig J**, **Millainathan S and Obermeyer Z** (2015) Prediction policy problems. *The American Economic Review 105*(5), 491–495. https://doi.org/10.1257/aer.p20151023

Kramer ADI, **Guillory JE and Hancock JT** (2014) Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks. *Proceedings of the National Academy of Sciences 111*(24), 8788–8790. https://doi.org/10.1073/pnas.1320040111

Kreuter F, **Ghani R and Lane J** (2019) Change through data: A data analytics training program for government employees. *Harvard Data Science Review 1*(2), 1–26. https://doi.org/10.1162/99608f92.ed353ae3

Krumm J (ed.) (2018) *Ubiquitous Computing Fundamentals*. New York: CRC Press.

Krumpal I (2013) determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity 47*(4), 2025–2047. https://doi.org/10.1007/s11135-011-9640-9

Kumar SA, **Madhumathi R**, **Chelliah PR**, **Tao L and Wang S** (2018) A novel digital twin-centric approach for driver intention prediction and traffic congestion avoidance. *Journal of Reliable Intelligent Environments 4*(4), 199–209. https://doi.org/10.1007/s40860-018-0069-y

Lazer D and Radford J (2017) Data Ex Machina: Introduction to big data. *Annual Review of Sociology 43*(1), 19–39. http://doi.org/10.1146/annurev-soc-060116-053457

Lazer DMJ, **Pentland A**, **Watts DJ**, **Aral S**, **Athey S**, **Contractor N**, **Freelon D**, **Gonzalez-Bailon S**, **King, G**, **Margetts H**, **Nelson A**, **Salganik MJ**, **Strohmaier M**, **Vespignani A and Wagn C** (2020) Computational social science: Obstacles and opportunities. *Science 369*, 1060–1062. https://doi.org/10.1126/science.aaz8170

LeCun Y, **Bengio Y and Hinton G** (2015) Deep learning. *Nature 521*, 436–444. https://doi.org/10.1038/nature14539

Lensvelt-Mulders GJ, **Hox JJ**, **Van der Heijden PG and Maas CJ** (2005) Meta-analysis of randomized response research: Thirty-five years of validation. *Sociological Methods & Research 33*(3), 319–348. http://doi.org/10.1177/0049124104268664

Liiv I (2021) *Understanding the data model*. In Liiv I (ed.), Behaviormetrics: Quantitative Approaches to Human Behavior Singapore: Springer, 1–13.

Liu L, **Silva EA**, **Wu C and Wang H** (2017) A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Computers, Environment and Urban Systems 65*, 113–125. https://doi.org/10.1016/j.compenvurbsys.2017.06.003

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

Liu M, **Han L**, **Xiong S**, **Qing L**, **Ji H and Peng Y** (2019) Large-scale street space quality evaluation based on deep learning over street view image. In *International Conference on Image and Graphics*, 690–701.

Luna DR, **Mayan JC**, **García MJ**, **Almerares AA and Househ M** (2014) Challenges and Potential Solutions for Big Data Implementations in Developing Countries. *Yearbook of Medical Informatics 9*(1), 36–41. https://doi.org/10.15265/IY-2014-0012

Mackey TK, **Kalyanam J**, **Katsuki T and Lanckriet G** (2017) Twitter-Based Detection of Illegal Online Sale of Prescription Opioid. *American Journal of Public Health 107*(12), 1910–1915. https://ajph.aphapublications.org/doi/abs/10.2105/AJPH.2017.303994

Mandalari AM, **Dubois DJ**, **Kolcun R**, **Paracha MT**, **Haddadi H and Choffnes D** (2021) Blocking without breaking: Identification and mitigation of non-essential IoT traffic. *arXiv* preprint 2105.05162. Available at https://doi.org/10.48550/arXiv.2105.05162

Marbouh D, **Simsekler MCE**, **Salah K**, **Jayaraman R and Ellahham S** (2022) Blockchain for Patient Safety: Use cases opportunities and open challenges. *Data 7*(12). https://doi.org/10.3390/data7120182.

McKelvey K, **DiGrazia J and Rojas F** (2014) twitter publics: how online political communities signaled electoral outcomes in the 2010 US house election. *Information, Communication & Society 17*(4), 436–450. http://doi.org/10.1080/1369118X.2014.892149

McQuinn A and Castro D (2019) *A Policymaker's Guide to Blockchain.* Information Technology and Innovation Foundation. Available at https://itif.org/publications/2019/04/30/policymakers-guide-blockchain/

Meijer A and Wessels M (2019) Predictive Policing: Review of Benefits and Drawbacks. *International Journal of Public Administration 42*(12), 1031–1039. http://doi.org/10.1080/01900692.2019.1575664

Mellado B, **Wu J**, **Kong JD**, **Bragazzi NL**, **Asgary A**, **Kawonga M**, **Choma N**, **Hayasi K**, **Lieberman B**, **Mathaha T**, **Mbada M**, **Ruan X**, **Stevenson F and Orbinski J** (2021) Leveraging artificial intelligence and big data to optimize COVID-19 Clinical public health and vaccination roll-out strategies in Africa. *International Journal of Environmental Research and Public Health 18* (15). https://doi.org/10.3390/ijerph18157890

Mergel I, **Rethemeyer RK and Isett K** (2016) Big data in public affairs. *Public Administration Review*, *76*(6), 928–937. https://doi.org/10.1111/puar.12625

Merry K and Bettinger P (2019) Smartphone GPS accuracy study in an urban environment. *PloS One 14*(7). https://doi.org/10.1371/journal.pone.0219890

Minoli D and Occhiogrosso B (2018) Blockchain mechanisms for IoT security. *Internet of Things*, 1–2, 1–13. https://doi.org/10.1016/j.iot.2018.05.002

Mirabelli G and Solina V (2020) Blockchain and agricultural supply chains traceability: Research trends and future challenges. *Procedia Manufacturing 42*, 414–421. https://doi.org/10.1016/j.promfg.2020.02.054

Morisette JT, **Cravens AE**, **Miller BW**, **Talbert M**, **Talbert C**, **Jarnevich C**, **Fink M**, **Decker K and Odell EA** (2017) Crossing boundaries in a collaborative modeling workspace. *Society & Natural Resources 30*(9), 1158–1167. http://doi.org/10.1080/08941920.2017.1290178

Mueller S and Pearl J (2023) Personalized decision making – A conceptual introduction. *Journal of Causal Inference 11*(1), 20220050. https://doi.org/10.1515/jci-2022-0050

Murphy J, **Link MW**, **Childs JH**, **Tesfaye CL**, **Dean E**, **Stern M**, **Pasek J**, **Cohen J**, **Callegaro M and Harwood P** (2014) Social media in public opinion research: Executive summary of the AAPOR task force on emerging technologies in public opinion research. *Public Opinion Quarterly 78*(4), 788–794. https://doi.org/10.1093/poq/nfu053

Mohammadi N and Taylor J (2021) Thinking fast and slow in disaster decision-making with Smart City Digital Twins, *Nature Computational Science 1*, 771–773. https://doi.org/10.1038/s43588-021-00174-0

Jalal N, **Alon I and Paltrinieri A** (2021) A bibliometric review of cryptocurrencies as a financial asset. *Technology Analysis & Strategic Management* 1–16. https://doi.org/10.1080/09537325.2021.1939001

Niederer SA, **Sacks MS**, **Girolami M and Willcox K** (2021) Scaling digital twins from the artisanal to the industrial. *Nature Computational Science 1*(5), 313–320. http://doi.org/10.1038/s43588-021-00072-5

Nikolic M and Bierlaire M (2017) Review of transportation mode detection approaches based on smartphone data. In *17th Swiss Transport Research Conference.* Monte Verità, Ascona: STRC.

Nordbotten S (2010) *The Use of Administrative Data in Official Statistics-Past, Present and Future: With Special Reference to the Nordic Countries.* Available at https://ssb.brage.unit.no/ssb-xmlui/bitstream/handle/11250/181409/Nordbotten_the%20use%20of%20administrative%20data_2010.pdf?sequence=1

Ojo A and Adebayo S (2017) Blockchain as a next generation government information infrastructure: A review of initiatives in D5 countries. In Adegboyega Ojo and Jeremy Millard (ed.), *Government 3.0 – Next Generation Government Technology Infrastructure and Services.* pp. 283–298.http://doi.org/10.1007/978-3-319-63743-3_11

Omar IA, **Jayaraman R**, **Salah K**, **Simsekler MCE**, **Yaqoob I and Ellahham S** (2020) Ensuring protocol compliance and data transparency in clinical trials using Blockchain smart contracts. *BMC Medical Research Methodology 20*(224). https://doi.org/10.1186/s12874-020-01109-5

Opara A, **Johng H**, **Hill T and Chung L** (2022) A framework for representing internet of things security and privacy policies and detecting potential problems. *In Proceedings of the ACM/SIGAPP Symposium on Applied Computing.* Association for Computing Machinery, 198–201.https://doi.org/10.1145/3477314.3508385

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

Pallmann P, **Bedding AW**, **Choodari-Oskooei B**, **Dimairo M**, **Flight L**, **Hampson LV**, **Holmes J**, **Mander AP**, **Odondi L**, **Sydes M**, **Villar SS**, **Wason JS**, **Weir CJ**, **Wheeler GM and Jaki T** (2018) Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Medicine 16*(1), 1–15. https://doi.org/10.1186/s12916-018-1017-7

Papyshev G and Yarime M (2021) Exploring city digital twins as policy tools: A task-based approach to generating synthetic data on urban mobility. *Data & Policy 3*, E16. https://doi.org/10.1017/dap.2021.17

Pantalona G, **Tsalakanidou F**, **Nikolopoulos S**, **Kompatsiaris I**, **Lombardo F**, **Norbiato D and Haberstock H** (2021) Decision support system for flood risk reduction policies: The case of a flood protection measure in the area of Vicenza. *Data & Policy 3*, E26. https://doi.org/10.1017/dap.2021.23

Phillips J, **Babcock R and Orbinski J** (2022) The digital response to COVID-19: Exploring the use of digital technology for information collection, dissemination, and social control in a global pandemic. *Journal of Business Continuity & Emergency Planning* https://doi.org/10.6084/m9.figshare.19950218.v3

Pottie GJ and Kaiser WJ (2000) Wireless integrated network sensors. *Communications of the ACM*, *43*(5), 51–58.

Purkayastha S and Braa J (2013) Big Data Analytics for developing countries—Using the Cloud for Operational BI in Health. *Electronic Journal of Information Systems in Developing Countries 59*(1). https://doi.org/10.1002/j.1681-4835.2013.tb00420.x

PWC (2019) Establishing blockchain policy: Strategies for the governance of distributed ledger technology ecosystems. Available at https://www.pwc.com/m1/en/publications/documents/establishing-blockchain-policy-pwc.pdf

Ratti C and Claudel M (2016) *The City of Tomorrow: Sensors, Networks, Hackers, and the Future of Urban Life*. Yale University Press.

Rao JNK (2021) On making valid inferences by integrating data from surveys and other sources. *Sankhya B: The Indian Journal of Statistics 83*(1), 242–272. https://doi.org/10.1007/s13571-020-00227-w

Remotti L (2021) IoT innovation clusters in Europe and the case for public policy. *Data & Policy 3*, E25. https://doi.org/10.1017/dap.2021.16

Renieris EM (2023) *The Best Way to Govern AI? Emulate It.* Centre for International Governance Innovation. Available at https://www.cigionline.org/articles/the-best-way-to-govern-ai-emulate-it/ (accessed 10 August 2023).

Ruggles S, **Fitch C**, **Magnuson D and Schroeder J** (2019). Differential privacy and census data: Implications for social and economic research. *AEA Papers and Proceedings 109*, 403–08.

Rundle AG, **Bader MDM**, **Richards CA**, **Neckerman KM and Teitler JO** (2011) Using Google Street View to Audit Neighborhood Environments. *American Journal of Preventive Medicine 40*(1), 94–100. https://doi.org/10.1016/j.amepre.2010.09.034

Salah K, **Rehman MHU**, **Nizamuddin N**, **Al-Fuqaha A** (2019) Blockchain for AI: Review and open research challenges. *IEEE Access 7*, 10127–10149. https://doi.org/10.1109/ACCESS.2018.2890507

Salem F (2017) Social media and the internet of things towards data-driven policymaking in the Arab world: potential, limits and concerns. *The Arab Social Media Report*, Dubai: MBR School of Government, 7.

Salganik MJ (2019) *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.

SAS (2014) Government workforce in focus: Closing the data and analytics skills gap. Washington, DC: SAS. Available at https://onlinebusiness.american.edu/wp-content/uploads/sites/69/2021/06/Closing-the-Data-and-Analytics-Skills-Gap-Research-Brief.pdf

Satyanarayanan M (2001) Pervasive computing: Vision and challenges. *IEEE Personal Communications 8*(4), 10–17. https://doi.org/10.1109/98.943998

Schweinfest S and Jansen R (2021) Data science and official statistics: Toward a new data culture. *Harvard Data Science Review* **3** (4)https://doi.org/10.1162/99608f92.c1237762

Sen I, **Floeck F**, **Weller K**, **Weiss B and Wagner C** (2019) A total error framework for digital traces of humans. *arXiv preprint 1907.08228.* Available at https://www.researchgate.net/publication/334603039_A_Total_Error_Framework_for_Digital_Traces_of_Humans

Singapore Land Authority (n.d.). Virtual Singapore. Available at: https://www.sla.gov.sg/articles/press-releases/2014/virtual-singapore-a-3d-city-model-platform-forknowledge-sharing-and-community-collaboration (accessed 19 February 2025)

Sharma M, **Joshi S**, **Kannan D**, **Govindan K**, **Singh R and Purohit HC** (2020) Internet of Things (IoT) adoption barriers of smart cities' waste management: An Indian context. *Journal of Cleaner Production 270*. https://doi.org/10.1016/j.jclepro.2020.122047

Singer E (2006) Introduction: Nonresponse bias in household surveys. *International Journal of Public Opinion Quarterly 70*(5), 637–45. http://doi.org/10.1093/poq/nfl034

Singleton AD, **Spielman S and Folch D** (2017) *Urban Analytics (Spatial Analytics and GIS)*. London: Sage.

Solman H, **Kirkegaard JK**, **Smits M and Van Vliet B** (2022) Digital twinning as an act of governance in the wind energy sector. *Environmental Science & Policy 127*, 272–279. https://doi.org/10.1016/j.envsci.2021.10.027

Stopher P, **FitzGerald C and Xu M** (2007) Assessing the accuracy of the sydney household travel survey with GPS. *Transportation 34*(6), 723–741.

Suominen A and Hajikhani A (2021) Research themes in big data analytics for policymaking: Insights from a mixed-methods systematic literature review. *Policy and Internet 13*(4), 464–484. https://doi.org/10.1002/poi3.258

Tanczer LM, **Brass I**, **Elsden M**, **Carr M and Blackstock J** (2019) *The United Kingdom's Emerging Internet of Things (IoT) Policy Landscape.*. In Ellis R and Mohan V (eds.), *Rewired*: *Cybersecurity Governance*. Hoboken, New Jersey: Wiley, pp. 37–56

Tao F and Qi Q (2019) Make more digital twins. *Nature 573*, 490–491. https://doi.org/10.1038/d41586-019-02849-1

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

Tapscott A and Tapscott D (2017) *How Blockchain Is Changing Finance.* Harvard Business Review Blog. Available at https://hbr.org/2017/03/how-blockchain-is-changing-finance

Taylor E and Gill M (2014) CCTV: reflections on its use, abuse and effectiveness. In *The Handbook of Security*, Springer, 705–726.

Thorpe HH (2023) ChatGPT is fun, but not an author. *Science 379*(6630), 313. https://doi.org/10.1126/science.adg7879

Tilak S, **Abu-Ghazaleh NB, and Heinzelman W** (2002) A taxonomy of wireless micro-sensor network models. *ACM SIGMOBILE Mobile Computing and Communications Review*, *6*(2), 28–36.

Tourangeau R and Yan T (2007) Sensitive Questions in Surveys. *Psychological Bulletin 133*(5), 859–883. https://psycnet.apa.org/doi/10.1037/0033-2909.133.5.859

Tsoi KKF, **Sung JJY**, **Lee HWY**, **Yiu KKL**, **Fung H and Wong SYS** (2021) The way forward after COVID-19 vaccination: vaccine passports with blockchain to protect personal privacy. *BMJ Innovations 7*, 337–341. https://doi.org/10.1136/bmjinnov-2021-000661

Tumasjan A, **O'Sprenger T**, **Sandner PG and Welpe IM** (2011) Election Forecasts with Twitter: How 140 Characters Reflect the Political Landscape. *Social Science Computer Review 29*(4), 402–418. https://doi.org/10.1177/0894439310386557

Ukil A, **Bandyopadhyay S and Pal A** (2014) IoT-privacy: To be private or not to be private. In 2014 *IEEE Conference on Computer Communications Workshops*, 123–124.

Ulibarri N (2018) Collaborative model development increases trust in and use of scientific information in environmental decision-making. *Environmental Science & Policy 82*, 136–142. https://doi.org/10.1016/j.envsci.2018.01.022

Upadhyay A, **Mukhuty S**, **Kumar V and Kazancoglu Y** (2021) Blockchain technology and the circular economy: Implications for sustainability and social responsibility. *Journal of Cleaner Production 293*. https://doi.org/10.1016/j.jclepro.2021.126130

van de Sanden S, **Willems K and Brengman M** (2019) In-Store Location-Based Marketing with Beacons: From Inflated Expectations to Smart Use in Retailing. *Journal of Marketing Management 35*(15–16), 1514–1541. https://doi.org/10.1080/0267257X.2019.1689154

Vaswani A, **Shazeer N**, **Parmar N**, **Uszkoreit J**, **Jones L**, **Gomez A**, **Kaiser L** (2017) Attention is All You Need. In *31st Conference on Neural Information Processing Systems.* Long Beach: NIPS

Veale M, **van Kleek M and Binns R** (2018) Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. New York, NY: Association for Computing Machinery, pp. 1–14. https://doi.org/10.1145/3173574.3174014

Vinuesa R and Sirmacek B (2021) Interpretable deep-learning models to help achieve the Sustainable Development Goals. *Nature Machine Intelligence 3*, 926. https://doi.org/10.1038/s42256-021-00414-y

Verhulst S, **Engin Z and Crowcroft J** (2019) Data & Policy: A new venue to study and explore policy–data interaction. *Data & Policy 1*, E1. https://doi.org/10.1017/dap.2019.2

Verhulst S (2021) Reimagining data responsibility: 10 new approaches toward a culture of trust in re-using data to address critical public needs. *Data & Policy 3*, E6. https://doi.org/10.1017/dap.2021.4

Vermesan O, **Friess P**, **Guillemin P**, **Sundmaeker H**, **Eisenhauer M**, **Moessner K**, **Gall F and Cousin P** (2022) Internet of things strategic research and innovation agenda. In *Internet of things* River Publishers, pp. 7–151.

Wager S and Athey S (2018) Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association 113*(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839

Wagner C, **Strohmaier M**, **Olteanu A**, **Kıcıman E**, **Contractor N and Eliassi-Rad T** (2021) Measuring algorithmically infused societies. *Nature 595*(7866), 197–204. https://doi.org/10.1038/s41586-021-03666-1

Wania A, **Kemper T**, **Tiede D and Zeil P** (2014) Mapping Recent Built-up Area Changes in the City of Harare with High Resolution Satellite Imagery. *Applied Geography 46*, 35–44. http://doi.org/10.1016/j.apgeog.2013.10.005

Wang J, **Xu C**, **Jiang J and Zhong R** (2022) Big data analytics for intelligent manufacturing systems: A review. *Journal of Manufacturing Systems 62*, 738–752. https://doi.org/10.1016/j.jmsy.2021.03.005

World Economic Forum (2023) Future of Jobs Report. Available at: https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf (accessed 15 August 2023)

Wolf J, **Oliveira M and Thompson M** (2003) Impact of Underreporting on Mileage and Travel Time Estimates: Results from Global Positioning System-Enhanced Household Travel Survey. *Transportation Research Record 1854*(1), 189–198. https://doi.org/10.3141/1854-21

Wong S, **Yeung JKW**, **Lau, YY and So J** (2021) Technical sustainability of cloud-based Blockchain integrated with machine learning for supply chain management. *Sustainability 13*, 8270. https://doi.org/10.3390/su13158270

Wright L and Davidson S (2020) How to tell the difference between a model and a digital twin. *Advanced Modeling and Simulation in Engineering Sciences 7*(13). https://doi.org/10.1186/s40323-020-00147-4

Wu X, **Brown KN and Sreenan CJ** (2013) Analysis of Smartphone User Mobility Traces for Opportunistic Data Collection in Wireless Sensor Networks. *Pervasive and Mobile Computing 9*(6), 881–891. https://doi.org/10.1109/TMC.2016.2595574

Yang Z, **Dai Z**, **Yang Y**, **Carbonell J**, **Salakhutdinov R and Le QV** (2019) XLNet: generalized autoregressive pretraining for language understanding. In *33rd Conference on Neural Information Processing Systems.* Vancouver: NeurIPS

Yang L, **Holtz D**, **Jaffe S**, **Suri S**, **Sinha S**, **Weston J**, **Joyce C**, **Shah N**, **Sherman K**, **Hecht B and Teevan J** (2021) The Effects of Remote Work on Collaboration Among Information Workers. *Nature Human Behaviour 6*, 43–54. https://doi.org/10.1038/s41562-021-01196-4

Yao H, **Rashidan S**, **Dong X**, **Hongyi D**, **Rosenthal RN and Wang F** (2021) Detection of Suicidality Among Opioid Users on Reddit: Machine Learning–Based Approach. *Journal of Medical Internet Research 22*(11). https://doi.org/10.2196/15293

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press

Ye Y, **Zeng W**, **Shen Q**, **Zhang X and Lu Y** (2019) The Visual Quality of Streets: A Human-Centred Continuous Measurement Based on Machine Learning Algorithms and Street View Images. *Environment and Planning B: Urban Analytics and City Science 46*(8), 1439–1457. https://doi.org/10.1177/2399808319828734

Yeoh P (2017) Regulatory issues in blockchain technology. *Journal of Financial Regulation and Compliance 25*(2), 196–208. https://doi.org/10.1108/JFRC-08-2016-0068

Zhang F, **Wu L**, **Zhu D and Liu Y** (2019) Social Sensing from Street-Level Imagery: A Case Study in Learning Spatio-Temporal Urban Mobility Patterns. *ISPRS Journal of Photogrammetry and Remote Sensing 153*, 48–58. http://doi.org/10.1016/j.isprsjprs.2019.04.017

Zutshi A, **Grilo A and Nodehi T** (2021) The value proposition of blockchain technologies and its impact on Digital Platforms. *Computers & Industrial Engineering 155*. https://doi.org/10.1016/j.cie.2021.107187

Cite this article: Asensio OI, Moore CE, Ulibarri N, Simsekler MCE, Lan T and Rivero G (2025). Data technologies and analytics for policy and governance: a landscape review. *Data & Policy*, 7: e25. doi:10.1017/dap.2024.49

https://doi.org/10.1017/dap.2024.49 Published online by Cambridge University Press